

**Manuscript version: Author's Accepted Manuscript**

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

**Persistent WRAP URL:**

<http://wrap.warwick.ac.uk/114490>

**How to cite:**

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk).

# Audio-Visual-Olfactory Resource Allocation for Tri-modal Virtual Environments

E. Doukakis, K. Debattista, T. Bashford-Rogers, A. Dhokia, A. Asadipour, A. Chalmers, and C. Harvey

**Abstract**—Virtual Environments (VEs) provide the opportunity to simulate a wide range of applications, from training to entertainment, in a safe and controlled manner. For applications which require realistic representations of real world environments, the VEs need to provide multiple, physically accurate sensory stimuli. However, simulating all the senses that comprise the human sensory system (HSS) is a task that requires significant computational resources. Since it is intractable to deliver all senses at the highest quality, we propose a resource distribution scheme in order to achieve an optimal perceptual experience within the given computational budgets. This paper investigates resource balancing for multi-modal scenarios composed of aural, visual and olfactory stimuli. Three experimental studies were conducted. The first experiment identified perceptual boundaries for olfactory computation. In the second experiment, participants ( $N = 25$ ) were asked, across a fixed number of budgets ( $M = 5$ ), to identify what they perceived to be the best visual, acoustic and olfactory stimulus quality for a given computational budget. Results demonstrate that participants tend to prioritize visual quality compared to other sensory stimuli. However, as the budget size is increased, users prefer a balanced distribution of resources with an increased preference for having smell impulses in the VE. Based on the collected data, a quality prediction model is proposed and its accuracy is validated against previously unused budgets and an untested scenario in a third and final experiment.

**Index Terms**—Multi-Modal, Cross-Modal, Tri-Modal, Sound, Graphics, Olfactory

## 1 INTRODUCTION

The existence of multiple stimuli in a VE is required for increasing immersion in many current and future applications. The inclusion of realistic olfactory delivery along with audio-visual stimuli is of key importance if VEs are to be used as genuine representations of real life scenarios [36]. The introduction of smell impulses increases the sense of presence in the virtual world and enhances the level of realism [16]. The significance of a tri-modal combination of smell, vision, and hearing might not be initially obvious but it affects how people think, feel and behave [40].

Accurate computation and delivery of multiple stimuli in high fidelity requires significant computing capability. Previous research has shown that humans cannot fully attend to all the incoming sensory stimuli in the real environment. Such cross-modal interaction phenomena and known limitations of the HSS have been utilized to reduce computational requirements [29] without the users being able to perceive any quality degradations. Furthermore, it is unclear how best to allocate computational resources in a VE. For example, if an improvement in computational performance of 50% becomes available, how would it be best to make use of that extra computational power. Is it best spent on improving visuals? Improving the audio? A smaller improvement to both? Adding smell? Resource allocation schemes have been recently proposed to describe distribution of available resources based on human subjective preferences [19]. However, previous work is limited to two modalities, in particular, audio and vision.

The work presented in this paper captures how human allocation preferences are adjusted given a budget of computational resources in a tri-modal VE set up. In order to establish how this computational budget can be distributed between the three senses, three experiments have been conducted. Experiment I was focused on understanding the perceptual limitations of olfactory computation, such that the compu-

tational requirements for olfactory delivery can be established. With this knowledge, Experiment II was conducted whereby participants allocated computational resources across visual, auditory and olfactory stimuli in order to identify a perceptually optimal load balancing of these resources from a given fixed budget. This was conducted at five different computational budgets for a number of scenarios. Based on the collected subjective data, a resource distribution model is proposed and evaluated on untested budget sizes and scenarios in Experiment III. Therefore, multi-sensory rendering pipelines can exploit such a model to direct resource allocation decisions in VEs.

The main contributions of this work are as follows:

- A psychophysics framework for estimating odour just noticeable difference (JND) thresholds for a range of smell concentration magnitudes.
- An experimental methodology for allocating resources in tri-modal VEs.
- Evidence that participants generally prefer to allocate resources for visual stimuli. However, as the budget increases the percentage devoted for aural and olfactory stimuli in the virtual scenario is increased significantly.
- A validated model capable of predicting resource allocation in systems where visual-aural and olfactory cues are intended for delivery.

## 2 BACKGROUND AND RELATED WORK

Simulation and delivery of multiple senses at the same time is considered crucial for ensuring a realistic experience and increase a user's overall level of immersion [21]. Applications of multi-sensory VEs range across different sectors of academia and industry. In reality, perceiving one sensory stimulus is quite rare in the physical world and many studies that investigate multiple senses, have found that the perceptual impact of one sense to the other can be quite significant [5, 20].

### 2.1 Audio-Visual Interactions

Increases in reported presence have been found in audio-visual VEs across a range of acoustic conditions [9, 27, 32]. Indeed, overall quality perception in audio-visual environments is mediated by the quality in each sense [44]. This showed that better audio increases the perception of quality in the visual domain whilst the opposite was also observed:

- Correspondence to E.Doukakis, stratosver2@gmail.com or K.Debattista, k.debattista@warwick.ac.uk
- E. Doukakis, K. Debattista, A. Dhokia, A. Asadipour and A. Chalmers are with University of Warwick
- T. Bashford-Rogers is with University of the West of England
- C. Harvey is with Birmingham City University

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxxx

better visuals decreased the perception of quality in the acoustic domain. This effect was shown to be practical by Moeck *et al.* by using hierarchical clustering of sound sources given congruent visual signals [34]. Perceptual interactions as triggered by sound cues are explored by Rocchesso [13]. Using sound effects, a series of human-human and human-object interactions are explained and validated. Preserving the level of presence in a VE has recently been examined by Grani *et al.* [8]. In their work, audio-visual attractors are used in an experimental study to quantify how users' attention is directed in a cave automatic virtual environment, avoiding gaps in presence. Sound has also been shown to influence perception in the spatial domain, such that congruent sounds can direct attention. This effect has been used in selective rendering models [26].

## 2.2 Olfactory-Visual Interactions

Olfactory-visual stimulation has been shown to increase presence in both generalized virtual environments and in targeted virtual environments when compared to a visual only condition [18, 35]. Supplementary to this, Munyan *et al.* [35], showed that when the olfaction condition was removed, this resulted in a disproportionate decrease in presence. Attentional changes have been observed when humans are presented with multi-sensory stimuli when compared to a visual only condition. Seo *et al.* [42] observed congruent objects impacting viewing time and deviating eye fixation. Seignuric *et al.* [41] investigated the influence of *a-priori* connections in between a scent and congruent visual stimuli on eye saccades and fixations, showing that congruent objects were explored faster in the presence of the odour. Chen *et al.* [15] performed a study to corroborate this effect and concluded that a multi-modal saliency map weighing both visual and olfactory inputs was required. Harvey *et al.* [25] showed that conventional image saliency maps can no longer be relied upon in the same way in olfactory-visual environments and demonstrated a validated model based on empirical findings.

## 2.3 Multisensory Integration

Burr and Allais have proposed a linear model for bimodal fusion in the audio-visual domain [6]. This suggests that weights control the bimodal information from the two senses:  $\hat{S} = w_A \hat{S}_A + w_V \hat{S}_V$ , where  $w_A$  and  $w_V$  scale the estimates for audio and vision respectively,  $\hat{S}_A$  and  $\hat{S}_V$ . Multisensory VEs are computationally demanding when considering the simulation of numerous senses [22]. It is however possible to balance computation to account for the weight that the human sensory system places on each sense. However, multisensory VEs have inherent perceptual affects that have to be understood [4], before these weights can be derived. In the study proposed by Doukakis *et al.* [19], the authors presented a method for resource allocation in bi-modal VEs, namely vision and hearing, based on human subjective preferences. In that experimental study, participants allocated a given budget of resources to improve the quality of the audio-visual stimuli. Based on the results, an estimation model is proposed and validated. Similar methods have been used by Slater *et al.* [14] to investigate the level of presence in VEs by conducting experiments where experienced participants at immersive system methodologies vary four possible graphical factors. In a subsequent experimental study by Skarbez *et al.* [12], participants could adjust a series of coherence factors to increase the level of the plausibility illusion, to match the perceptual experience they had in a highest coherence scenario. Results showed that participants prioritize improvements to the virtual body.

Ernst and Banks [11] investigated which of the senses of vision and haptics is more dominant using a maximum-likelihood estimation on the combined input of both sensory cues. Using the variances of each sense in height estimation, a maximum-likelihood integrator model is given and compared to human collected data in visual-haptic tasks. Azevedo *et al.* [7] considered how the senses of vision, hearing, olfactory and haptics are classified for measuring presence, focusing on outdoor VEs. Their results showed that the combined effect of haptics and hearing was considered more important than the typical VE stimuli of vision and hearing, dependent on scene and plausibility illusions.

In summary, there exists a large body of work that considers the permutations of senses and their respective influences on human perception and presence. Studies that consider multisensory integration are shown to be of benefit in VEs when resources are adapted based upon empirical findings. In the bi-modal case, this resource allocation has been quantified but beyond remains unexplored.

## 3 MOTIVATION AND OVERVIEW

In this work, we are interested in identifying how to best allocate computational resources across audio, vision and olfaction. While audio and vision are reasonably well understood and have been used in a significant number of cross-modal experiments [29], this is not the case with olfaction. Following the audio-visual approach of Doukakis *et al.* [19], our tri-modal model is built around permitting users to adjust the required computation for all the senses to fit within a given computational budget. The selection and adjustment of the aural and visual stimuli is based on the approach adopted by Doukakis *et al.* [19]. However, since the application of olfaction in virtual environments is less understood, our initial experiment (Experiment I, Section 4) seeks to identify a useful perceptual parameterization for olfactory stimuli. Experiment II (Section 5) then collects data for load balancing across the tri-modal stimuli at five budgets across three scenarios. Section 6 uses these results to develop a model for tri-modal resource allocation. Finally, Experiment III (Section 7) validates the model.

## 4 EXPERIMENT I: IDENTIFYING OLFACTORY PARAMETERS

This section describes the olfactory simulation and its parametrisation. Olfaction will be parameterized by the mesh size (as we are using a finite element solver) as it effects the convergence rate of our olfactory simulation using computational fluid dynamics (CFD). An experimental framework is then described for estimating JNDs in a range of concentration magnitudes. The JND threshold estimation allows us to assess whether the recorded concentration curves for different mesh discretization levels are perceptually equivalent. As shall be shown, simulations across a wide range of mesh parameterizations (from 1K to 1M) are not perceptually noticeable by human observers.

### 4.1 Simulation

This section presents the framework used for simulating smell transport in four test VEs, as shown in Figure 1, using CFD. Odour concentration is captured at virtual probes during the simulation stage across a range of different quality meshes. In this work, the considered mixture is composed by the smell of citral in air and the objective is to estimate the concentration of citral during the transport of the mixture.

All the CFD simulations were implemented using the *Eulerian* approach, i.e. compute the variables of interest over time at the centres of the control volumes (CVs) that compose the domain's computational mesh. This procedure requires two steps. Firstly, it includes the discretisation and solution of the Navier-Stokes equations that govern the transport of the air-odour mixture and compute its momentum ( $\vec{U}$ ), pressure ( $p$ ) and density ( $\rho$ ). Secondly, coupling the governing equations with a transport differential equation for estimating the concentration of the odour, denoted as  $C$ , at every node of the mesh. This equation is of the general form:

$$\underbrace{\frac{\partial \rho C}{\partial t}}_{\text{temporal term}} + \underbrace{\nabla \cdot (\rho \vec{U} C)}_{\text{convection term}} - \underbrace{\nabla \cdot \vec{J}}_{\text{diffusion term}} = \underbrace{S_C(C)}_{\text{source term}}, \quad (1)$$

where the quantity  $\vec{J} = \rho D_{ca} \nabla C$  is the mass diffusion flux of the odour and  $D_{ca} = 8.23 \times 10^{-5} \text{ m}^2/\text{s}$  is the diffusion coefficient of citral into the air [33]. The term  $S_C(C)$  describes the effect of body forces and is used to model the effect of the Earth's gravitational field. The temporal term and the terms that model the physical processes of convection and diffusion are underlined in Equation 1. As is the case with every odorous gas, the flow transport of the citral-air mixture is highly turbulent. Turbulence effects were implemented using a realisable  $k - \epsilon$  model [45].

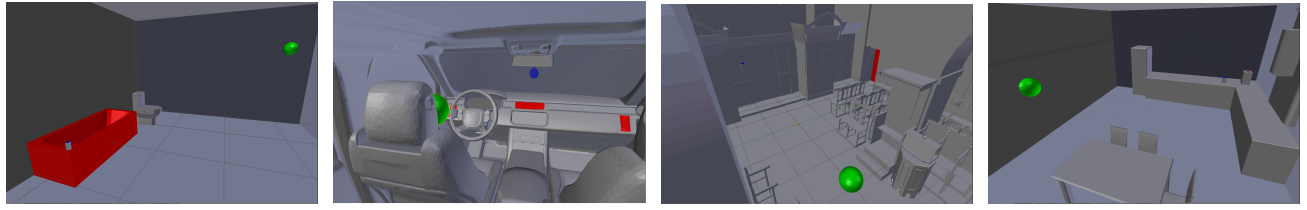


Fig. 1. The boundary conditions of the VEs used at this experimental study, From left to right: Bathroom interior, Car, Kitti, Kitchen. The blue painted patches represent the odour inlets while the red patches are used for developing convection effects. The green coloured spheres represent the virtual probes for reading concentration values.

The domain's mesh granularity affects the solution's convergence rate while it can increase numerical stability during the simulation [39]. Coarse meshes yield concentration solutions with over- and undershoots because of the non-smooth transition of the odour-air mixture across the CVs. We considered four different mesh versions of successively higher number of CVs for every one of the three scenarios. These are meshes with 1K, 10K, 100K and 1 million CVs. The coarsest mesh was refined near the surfaces and the boundaries of the VE for better accuracy while every successive refinement was uniform across the domain so as to approximately preserve the initial distribution of CVs in the boundaries. Mesh sizes with an order of magnitude change in the number of CVs were selected in order to study how odour concentration at probe locations changes between large spatial discretisation steps starting from very coarse up to excessively high refinement levels.

#### 4.1.1 Application to VEs

Simulation of smell propagation was considered in four different VEs. These are depicted in Figure 1. All the physical quantities used as boundary conditions were obtained by measuring concentration and flow rate with a photoionization detector and a flow meter respectively in the real places that were used for creating the VEs. The final values resulted through averaging of the results collected over 10 repeated measurements. These scenarios were chosen because the convection process occurs in a different way in each scenario, and therefore the smell will be dispersed differently. In the Bathroom, temperature differences between the hot bath (45 °C) and the environment (15 °C) cause the smell-air mixture to circulate in the room. In the car scenario, convection is simply created due to the air flow coming through the vents at a constant temperature of 25 °C. In the Kitti scenario, convection occurs due to the air coming through the three doors. In the Kitchen scenario, convection effects are introduced through the temperature gradient between the hot kettle (50 °C) and the cold environment (15 °C). Pressure was assumed to be the same and equal to atmospheric pressure (101.325 Pa) for all three VEs. Figure 1 depicts the smell inlet boundaries and the patches used for developing convection effects in the VEs. The same figure also shows the locations of the virtual probes in the VEs.

#### 4.1.2 Smell concentration results

Smell concentration values were recorded at the virtual probe using a sampling frequency of 4Hz. This sampling rate was chosen based on the rate many photoionisation devices record odour concentration values in real environments [23].

For all the VEs, smell propagation was simulated for 1,800 seconds of virtual time to see how concentration evolves at the sensor location for a sufficient amount of time. Figure 2 shows how concentration changes over time for the four different spatial discretisation levels. For coarse meshes the transition of the air-citral mixture near the sensor position is more abrupt therefore fluctuations are captured in the concentration function. These fluctuations are progressively eliminated as the mesh is refined. As can be seen from the graphs in Figure 2, the concentration curves are not very distant from each other. Specifically, the maximum difference between the coarsest (1K) and the finest meshes (1M) was 1.98 ppm for the Bathroom scenario, 1.623 ppm for the Car scenario and 1.345 ppm for the Kitti scenario. Furthermore, the curves

stabilise as the VE becomes saturated from the emitted odour. After this equilibrium point, concentration does not change significantly. The equilibrium point has not been fully achieved for the Kitti scenario because of the large volume of the room (560.64 m<sup>3</sup>) and the location of the sensor.

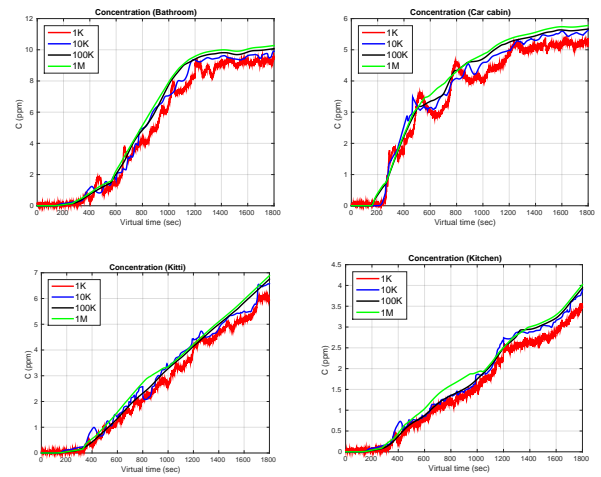


Fig. 2. Concentration results measured for different mesh discretisation levels at the virtual probe locations. Top-left: Bathroom, Top-right: Car, Bottom-Left: Kitti, Bottom-Right: Kitchen.

## 4.2 Method

In this study the experimental methodology used for estimating the JND threshold was the two interval forced choice (2IFC) method [30]. At every trial two olfactory stimuli were presented with an interstimulus time interval (ITI) between them. The first is a constant concentration stimulus, known as a pedestal, while the second is a varying stimulus randomly selected from a predefined set of concentration magnitudes. Ten trials were implemented for each comparison pair (pedestal, varying) for the different values of the varying stimulus. The presentation order of the two smell impulses (pedestal-varying or varying-pedestal) was also randomized across the experimental trials to avoid any potential ordering bias.

The JND threshold is the magnitude that needs to be added to the pedestal so as the resulting olfactory stimulation has a magnitude that is perceptually more intense than the pedestal in 75% of the total trials [30]. During each trial, the two stimuli were presented with an ITI interval of 4 seconds while the stimulus exposure duration (SED) was 0.5 seconds for each of the two delivered bursts. A short SED was selected to avoid user adaptation to the released smell while the ITI was sufficiently long to permit the participants to rest before inhaling the second stimulus.

In this work, a computer-controlled olfactory display was used to deliver the smell impulses at the two SEDs [17]. This device provides two independent channels that can release smell impulses synchronously



while it contains two digital mass flow controllers (DMFCs) for adjusting the flow rate of each channel as a proportion of the maximum flow rate provided which is 1000 ml/min. For our olfactory display, the suggested operating range for guaranteed concentration accuracy of the delivered stimulus is  $I = [C_0, C_{\max}] = [1.2, 11.2]$ , parts per million (ppm) for a SED time of 0.5 sec. All the participants were able to detect the lemon like odour when they sniffed  $C_0$ , thus, the whole range of stimuli contained in  $I$  were higher than the absolute detection threshold.

As there was no previous knowledge for the existence of any JND threshold in the range  $I$ , starting with  $C_0$  as the pedestal, pairwise comparisons of  $C_0$  with a number of varying stimuli that uniformly discretize  $I$  were planned until the average performance among all the participants would reach a proportion of 75%. For a step  $s$  the interval  $I$  was subdivided to three mutually exclusive subsets:

$$\begin{aligned} I_1 &= [1.2 + s, 1.2 + 2s, \dots, 1.2 + 8s] \\ I_2 &= [4.8 + s, 4.8 + 2s, \dots, 4.8 + 8s] \\ I_3 &= [8 + s, 8 + 2s, \dots, 8 + 8s]. \end{aligned} \quad (2)$$

A step size  $s$  equal to 0.4 was selected so as to keep a reasonable number of varying stimuli for each of the three subintervals and be able to span the whole range in the extreme case that no perceptually different magnitude existed between  $C_0$  and any of the stimuli in  $I_1, I_2, I_3$ . Starting with comparisons of the pedestal  $C_0$  with the stimuli in interval  $I_1$ , if there was no stimulus that triggers a perceptual difference, denoted as  $C_1$  the experimentation would be carrying on in a subsequent experimental session for the same pedestal and the stimuli with magnitudes in the interval  $I_2$  until the first stimulus that triggers a perceptual difference. If there was no such stimulus, the pedestal would be compared with stimuli in  $I_3$ . In case  $C_1$  could be found, there is no need to continue experimenting for the rest of the stimuli in any of the  $I_2$  and/or  $I_3$  as it was necessary to find perceptual differences starting from  $C_1$  as the new pedestal.

The first time that a JND threshold was found a new experimental session was initiated by adjusting the step for spanning the rest of the interval  $I$  using Weber's law. This law has been applied in the olfactory domain [3] and states that the ratio of the JND to the pedestal stimulus is constant. Specifically:

$$\frac{\text{JND}}{C_p} = \frac{C_f - C_p}{C_p} = k, \quad (3)$$

where  $C_p$  is the pedestal and  $C_f$  the final intensity stimulus that triggers a perceptual difference. The constant  $k$  can be obtained from the first JND, denoted as  $\text{JND}_1$ , and the first pedestal  $C_0 = 1.20$  ppm. Using the constant  $k$ , we can obtain an estimation of the final stimulus that elicits a perceptual difference starting from  $C_1$  as the new pedestal and applying again Weber's law. In case the estimation is not in the interval  $I$ , no further experimentation is required as it cannot be delivered by the olfactory display. If not, a subinterval with a new discretization step is derived and a new experimental session is initiated. The new subinterval includes Weber's law estimation as the potential stimulus that elicits perceptual difference. The new step is computed so as the new interval does not exceed 10 varying concentration magnitudes.

This adaptive method allowed us to speed up the process of finding a new stimulus that triggers a perceptual difference without comparing intermediate pairs that would be perceptually similar using the initial small step of 0.4. The derivation of a new step enabled avoiding unnecessary experimental trials. Using this methodology, two JND thresholds were found in the operating range  $I$ . The first one was included in  $I_1$  while the second one was found after one step adjustment using an interval of 10 varying stimuli.

#### 4.2.1 Materials

For this study a 15.6" laptop computer was used for controlling the olfactory display and the graphical user interface (GUI). The apparatus is shown schematically in Figure 3. Two tubes arrived near the nose from the olfactometer, one that delivered smell bursts at every experimental trial and one air control channel that was evacuating the old

smell stimuli by releasing humidified air at an exact flow rate of 1000 mL/min during the ITI.

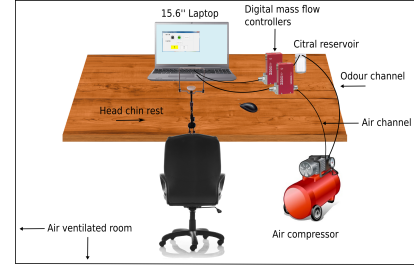


Fig. 3. Experimental set up for Experiment I.

The distance of the nose away from the smell delivery point was kept constant at 7 cm using a chin rest located in front of the laptop. An air compressor located under the table was used to provide two air channels for the experiment at a constant pressure of 2 Bars. The first channel was aromatized by passing through a smell reservoir where citral in liquid form was stored. Both the aromatized and the control channel were controlled using the DMFCs which were connected through USB ports on the laptop.

The odour of citral was used as the delivered olfactory stimulus in this experimental study. This is a frequently encountered volatile organic compound (VOC) reminiscent of the smell of lemon. Citral has a low detection threshold that ranges between 25 to 350 ppb [2]. All the participants were able to detect it and no participant reported the smell as unpleasant. The range  $I$  contained intensity stimulations that varied from faint up to more intense smell impulses without being irritating for the people who smelled them. The experimentation procedure was conducted under normal indoor lighting conditions in a room where air was constantly ventilated during the 4 days of conducting the experiment.

#### 4.2.2 Participants and Procedure

Two groups of 10 participants (11m, 9f, age range 19 to 35,  $\mu = 31.6$ ) from various academic and working disciplines. The average age of all the participants was 31.6 years old. No participant reported olfactory deficits or temporary/permanent anosmia problems.

For each trial, every participant had to smell the two bursts and then decide which one was the most intense by pressing one of the GUI buttons provided. During the tutorial session, pairs of ( $C_0, C_0$ ) stimuli were presented and the participant was asked if the smell was familiar or not. All the 20 participants were able to recognize the delivered odour. Participants were presented smells with the GUI providing context of the order and the participant could select the most intense stimulus from the GUI. Subsequently, the participant was permitted to select next, at their leisure, to continue onto the next trial. In the first phase each participant conducted eight trials and in the second they conducted ten trials.

#### 4.2.3 JND Results

Two JNDs were found in the operating range  $I$  provided by the olfactory display at two different experimentation phases. The first estimated stimulus that elicited a perceptual difference had magnitude in the subinterval  $I_1$  while the second was obtained using one step adjustment with the adaptive method described previously.

Phase 1 started with a pedestal stimulus  $C_0 = 1.20$  ppm while there were 8 varying stimuli to compare it with (see  $I_1$  interval at Equation 2). Figure 4 depicts the proportion of correct answers across all 10 participants participated in this phase and the Logistic Psychometric function (PF) fitted to the data. This function was preferred over others as it most accurately fitted the obtained data using deviance criterion as a model comparison metric [30]. For the 2IFC method, the stimulus that corresponds to 75% of correct answers is given through the PF and corresponds to the magnitude  $C_1 = 3.50$  ppm. Therefore, the first JND threshold is  $\text{JND}_1 = C_1 - C_0 = 2.30$  ppm.

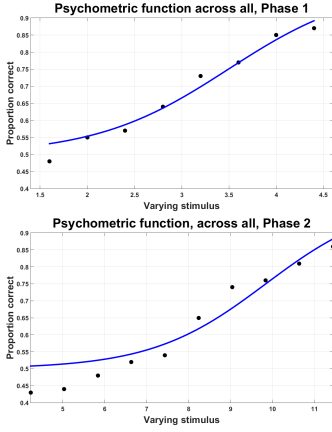


Fig. 4. Proportion of correct answers across all participants in Phase 1 (Top) and Phase 2 (Bottom). The blue colored curve is the Logistic Psychometric function fitted to the data in both cases.

The slope of the PF expresses how participants' performance changes with increasing concentration magnitudes. The estimate for this value was  $S_1 = 0.14$  units of performance per ppm. The standard error for the PF estimates for  $C_1$  and  $S_1$  were obtained based on a parametric bootstrap analysis for  $N = 1000$  generated data sets using the collected data of phase 1. For all the generated data collections the parameters  $C_1$  and  $S_1$  were estimated and their standard deviation was computed. These were  $SD_{C_1} = 0.12$  ppm and  $SD_{S_1} = 0.25$  ppm.

The goodness of fit for the selected PF was also computed to check if the selected PF sufficiently fits the data. The p-value for deviance found  $P_d = 0.69 > 0.05$  meaning that the Logistic (L) PF is a representative statistical model for the given data. The same criterion with Weibul (W), Gumbel (G), Cumulative Normal (CN) and Hyperbolic Secant (HS) PFs resulted in an acceptable  $P_d > 0.05$  but lower than the result given by the Logistic PF. The deviance values are given in Table 1.

The Weber constant for the results of phase 1 was  $k = 1.91$ . Using the  $C_1$  as the new pedestal, Weber's law (see formula 3) estimates that the new stimulus that elicits a perceptual difference is approximately 10.20 ppm. Using this outcome the step was adjusted as  $s = 0.8$  while a new experimental phase with a set of 10 varying stimuli was initiated. This set is:

$$I_a = \{C_1 + s, C_1 + 2s, \dots, 11.2\} \quad (4)$$

The last varying stimulus was 11.50 ppm which is out of the olfactory display's operating range, therefore it was replaced with the interval's maximum. The proportion of correct answers and the PF function are given in Figure 4. The stimulus that elicited a perceptual difference was found to be  $C_1 = 10.13$  ppm meaning that the second JND threshold is  $JND_2 = C_2 - C_1 = 6.63$  ppm. This estimation is relatively close to Weber's law prediction and the small difference is attributed due to experimentation error. The slope of the new PF is  $S_2 = 0.07$  units of performance per ppm. The standard error for both estimations is  $SD_{C_2} = 0.23$  ppm and  $SD_{S_2} = 0.13$  ppm using the same number of generated data sets as in phase 1 for the parametric bootstrap. The deviance p-value for assessing goodness of fit was  $P_d = 0.34 > 0.05$ , higher than all the other PFs tested (see Table 1).

	L	W	G	CN	HS
<b>Phase 1</b>	0.6949	0.581	0.579	0.631	0.629
<b>Phase 2</b>	0.34	0.301	0.293	0.318	0.325

Table 1. Deviance p-values for the different statistical models tested for goodness of fit for each of the two experimental phases.

### 4.3 Discussion

Smell transport simulation for different spatial discretisation levels and the JND experimental study yielded a number of potentially interesting findings. The experimental results clearly revealed that the human olfactory system (HOS) is relatively insensitive at small intensity variations. This result is confirmed from the high JND thresholds found at the whole operating range provided by the available hardware. The low standard deviations  $SD_{C_1}$  and  $SD_{C_2}$  of the estimates  $C_1$  and  $C_2$  that elicited perceptual differences show that the majority of the participants had difficulty in perceiving small intensity differences between the pedestal and varying stimuli. For olfactory stimulations of higher concentration the JND threshold is higher meaning that intensity variations are much more difficult to detect.

Although fine spatial discretisation levels of the computational domain can provide higher numerical accuracy in many applications, for olfaction, the JND threshold is clearly higher than the difference in odour intensity between the two extreme mesh sizes at every time point simulated. Accurate spatial discretisation of the computational domain in CFD simulations does not, on average, elicit quality improvements that can be consciously perceived by the users, hence the choice is made for olfaction to be a binary (on/off) choice in Experiment II.

## 5 EXPERIMENT II: TRI-MODAL RESOURCE ALLOCATION

This section describes the details of this experiment including experimental layout, material preparation, procedure and participants.

### 5.1 Design

The experimental design extends the design of Doukakis et al. [19], whereby participants are asked to allocate a given budget of resources by adjusting the quality of the displayed visual and auditory stimuli, to now include the delivery of smell impulses in a binary (on/off) fashion. The reasoning of on/off smell was based on the results of Experiment I. It is well known that the operation of the human olfactory mechanism is not fully understood [24, 38] and the evaluation of the incoming olfactory cues is highly subjective and based on previous experiences of the perceived scent [28, 40]. To avoid any bias of the quality levels due to these reasons, a two level scale that defines smell on or off is followed.

In this experiment, the audio-visual quality levels can be adjusted using the interlinked sliders of a GUI whilst two more controls are included for turning the smell stimulus on or off respectively. Physically-based simulations were used for the visual and aural stimuli whilst smell transport simulations were implemented for smell (see Section 4). The assigned budget allows the users to make audio-visual quality improvements and the cost of these improvements is deducted from the available budget. When the smell option is set to on, a percentage of the given budget is immediately reserved and the rest is given for audio-visual improvements.

The experiment was implemented with five distinct budgets across three different scenarios. One more scenario was used for training before the main experimental session. The experimental design is within participants and each participant was requested to make a judgement regarding the best perceived quality of all the possible combinations. The presentation order of the combinations was randomized to avoid any potential bias. In the rest of this section, the quality metrics for vision, audio and olfactory are explained. These metrics are used to derive quality levels for the three senses. The computation and metric selection for audio-visual stimuli was based on previous work [19] and a short summary of these metrics is presented here for completeness.

#### 5.1.1 Visual

In the visual domain, quality is adjusted using image resolution. Resolution is a standardized metric and it can be abstracted from the underlying algorithm used for the image computation. Other factors that can adjust image quality (samples per pixel, textures, etc.) are kept constant to avoid an exponential growth of the different possibilities. In this work, 240 images were computed using path tracing [10] starting from as low as  $16 \times 9$  up to the highest resolution at Ultra HD (UHD)

(3840 × 2160). The normalized computational cost for all the images in this sequence can be given as:

$$C_k^{N_V} = \left(\frac{k}{240}\right)^2, \quad k = 1, 2, \dots, 240. \quad (5)$$

As the visual differences between successive images can be very small, the original sequence was filtered using the High Dynamic Range Visible Difference Predictor (HDR-VDP) model [37] to discard those pairs that elicit the same perceptual response to the average user. The final sequence included 80 perceptually distinguishable images.

### 5.1.2 Audio

In the auditory domain, quality was adjusted using the sampling frequency of the computed room impulse response (RIR). Again a ray tracing approach [43] was used in the computation of the RIRs reaching a maximum sampling rate of 352,800 Hz. Similarly to the visual domain, 240 RIRs were computed at 240 different sampling rates. Every RIR was convoluted with an anechoic sound stream that contained the audio context of every scenario. The normalized computational times of every RIR can be estimated as:

$$C_{f_k}^{N_A} \approx \frac{f_k}{f_{240}}, \quad k = 1, 2, \dots, 240, \quad (6)$$

where  $f_k < f_{240}$ , and  $f_{240} = 352800$  Hz the maximum sampling rate. Based on the frequency JND distribution in the auditory domain [31], a filtering of the RIRs gives 80 log-spaced RIRs that clearly elicit sound differences after the convolution with the anechoic stream.

### 5.1.3 Olfactory

In the olfactory domain, a similar technique was followed for defining normalized costs for the olfactory cues. This technique is based on the perceptual properties found in Experiment I. The theoretical cost for an olfactory cue is estimated using the concentration results over virtual time found during the smell transport simulations. Specifically, for a given scenario, denoted by  $S$ , and a mesh size, denoted by  $M$ , the notation  $VT_M^S$  denotes the virtual time moment when the simulation reaches the concentration  $C_1 = 3.49$  ppm at the probe location for the chosen mesh and scenario. The concentration level  $C_1$  is the first olfactory stimulus that elicits a perceptual difference starting from concentration  $C_0 = 1.20$  ppm and was preferred as it is a medium intensity olfactory stimulus without being too faint or strong when sniffed. Using the notation  $PT_M^S$  for denoting the physical computation time needed for computing all the intermediate virtual times including  $VT_M^S$ , the normalized cost for smell is defined as:

$$A_M^S = \frac{PT_M^S}{PT_{1M}^S}, \quad (7)$$

where the subscript “1M” indicates the highest mesh quality used in the smell transport simulations (1 million CVs). The theoretical cost  $A_M^S$  is always between  $[0, 1]$  while it is an increasing function of the mesh size, i.e.  $A_{M_1}^S \leq A_{M_2}^S$  for meshes  $S_1, S_2$  where  $S_1$  has less CVs than  $S_2$ .

### 5.1.4 Tri-modal cost interactions

The choice of normalized cost functions in the audio, visual and olfactory domains, allows to investigate which of the three senses is more important for a given budget size without considering the problem of large differences in physical computation time for rendering an image, RIR or simulating an odour impulse in the VE. Also, normalisation of the costs makes the experimental results independent of the algorithmic strategy selected for computing the three stimuli. In this study, the total budget size is always distributed among the costs of a visual, auditory and an olfactory stimulus (when on). The budget sizes used in this study are given in Table 2. Note that  $B_5$  has fewer levels as the lowest qualities would not be available for this budget since the overall budget is higher without increasing the maximum qualities available for both audio and visuals.

Table 2. Theoretical budgets used in this experimental study along with their notation letters and the number of quality levels available when a budget is applied.

Budget letter	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	B <sub>5</sub>
Budget	0.0625	0.11	0.25	1	1.12
Total Number of Levels	28	38	48	80	48



Fig. 5. Scenarios of this experimental study. From Left to Right and Top to Bottom: Bathroom, Car, Kitchen, Kitti. The circular areas show the smell source.

The wide range of budget sizes ( $B_1$ - $B_5$ ) allows to investigate users' allocation preferences both with and without budget constraints. Using equation 7, and  $M = 1K$ , the normalized costs for the smell impulse can be computed for all four scenarios of interest. The theoretical cost for smell was estimated using the physical computation times obtained from the coarsest mesh size, therefore, the selection of 1K is justified from the findings of Experiment I, where the selection of the mesh size for simulating smell transport has no perceptual effect to the user.

Table 3. Theoretical costs for smell across the four scenarios selected. The first row shows the virtual time (sec) required to reach concentration  $C_1$ . The second row shows physical computation times (sec) for the coarsest mesh (1K). The third row shows physical computation times (sec) for the finest mesh (1M). The last row gives theoretical costs for smell as ratios of the second and third row.

Scenario	Bathroom	Car	Kitchen	Kitti
$VT_{1K}^S$	646	458	1685	1199
$PT_{1K}^S$	229.2	123.6	718.8	481.3
$PT_{1M}^S$	6222.5	4164.7	17900.3	15228.9
$A_{1K}^S$	0.037	0.029	0.040	0.032

## 5.2 Materials

For this study, two monitors (First: 28" Samsung U28D590 ultra HD LED monitor used for visual stimuli, Second: Dell UltraSharp 19" used for the GUI) and a set of headphones (Sennheiser HD 380 pro) were used for conducting this experiment. In addition, the olfactory display used of the preliminary experimental study was used for the delivery of the smell impulses. All the experimental trials were conducted in a silent room with constant ventilation. The same four scenarios used for smell transport simulations in Experiment I were used. The smell of citral, was congruent to all of them. Figure 5 depicts the experimental scenarios and zooms in the source of smell used for the simulation of odour transport. From the user's perspective, at least one object could have been the source of smell and no information was given to the participants regarding the source of the odour. A GUI application was used for carrying out the experimental study, see Figure 6.

The user becomes aware of the constraint in the budget as the red colored stripes increase in length and the number of available levels



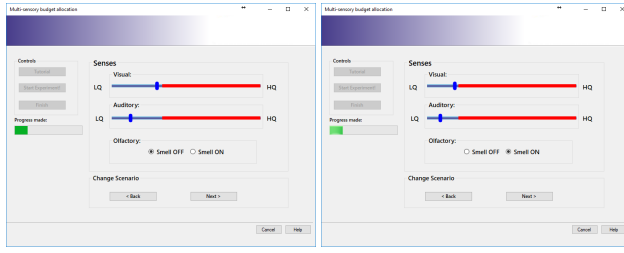


Fig. 6. Snapshots of the software used in the experimental study. Left: An instance of the budget size  $B_2$  when the “Smell OFF” control is activated. Right: An instance of the budget size  $B_2$  when the “Smell ON” control is activated.

for audio-visual quality improvements are reduced. When the control for smell was turned on, a smell impulse of 1 second was delivered to the user at close vicinity to the nose. The delivery of smell bursts instead of a continuous flow was preferred so as to avoid user’s fast adaptation to the released odour. A control channel was also used to release air bursts of 1,000 ml/min flow rate and convect the smell bursts away from participant’s location. Five air burst of frequency 400 ms were released when the user was pressing the “Next” or the “Smell OFF” buttons to go to the next trial or choose not to allocate resources when the “Smell ON” was enabled respectively. The delivery of the audio and visual stimuli was delayed by 240 ms when the smell was set on in the experiment to compensate the olfactory display’s onset time. The objective was to deliver all three sensory stimuli simultaneously without the participants perceiving temporal delays.

The use of the air channel was very important as it didn’t allow adaptation of the user to the smell cue. The air channel intended to disperse the smell plume quickly at the ventilation system which was right behind and above participant’s head. Also in situations where participants wanted to try the smell cue and disable it immediately, the air channel assisted to remove the residues of the smell out immediately in an attempt to disable the sensory cue as quickly as possible.

At the beginning of each experimental trial the two sliders are positioned at the start of the slider bars which correspond to a “null” stimulus. The “null” stimulus configuration includes the delivery of a grey image and a silent track while the smell control is pre-selected to be “OFF”. The initial selection of the grey image aims to neutralize participants’ eyes and is suggested by [1]. When budget  $B_5$  is available for resource allocation, the audio and visual thumbs start from a medium quality level that has cost equal to  $B_5 - B_4$ . The initial configuration of the thumbs allows the participants to explore all the available quality levels before deciding which quality level is desired. The alternative option to start the thumbs at random positions was not preferred to avoid biasing the participant with a thumb configuration which does not represent his/her actual preferences. At the beginning of each trial, the two thumbs are independent of each other until the sum of the theoretical costs for the three sensory stimuli exceeds the budget given for the trial. After the first attempt to exceed the budget, the thumb controls become dependent. At any point, if the user activates the smell impulse, the GUI sliders adjust to the new lower budget.

The transition from independent to dependent slider controls for adjusting audio-visual quality changes to either the senses of vision or hearing the quality of the other stimulus so as the budget to remain always constant and equal to the current budget size. The addition of the olfactory impulses does not disrupt this mechanism as the cost for a smell impulse is immediately allocated after the control is turned on.

### 5.3 Participants and Procedure

A total of 25 participants (13m, 12f, age range: 23 to 46,  $\mu = 36.2$ ) volunteered. All of them had normal or corrected to normal vision. None of them reported hearing or temporary/permanent smell problems. All the participants recruited for this experiment had no participation in Experiment I or were aware of its purpose.

Every experimental session was initiated with a tutorial where the

participants had the chance to familiarize with the task of allocating resources. The Kitti scenario was used for training with all possible budgets. During the tutorial, the experimenter had the chance to explain the objectives and purpose of the study to the participants for about 15 minutes before the main session. The participants were asked to judge and form the best multi-sensory experience using the controls of the GUI.

### 5.4 Results

The analysis of the results is divided into two parts for better clarity. The first part includes an ANOVA via a 3 (*scenario*)  $\times$  5 (*budget*) factorial design for studying the effect of the independent variables on participants’ decision to enable the smell impulse on and the second part describes a multivariate analysis of variance (MANOVA) for examining the effect of the two independent variables on the visual and auditory allocation preferences.

#### 5.4.1 Analysis of the smell preferences

Figure 7 depicts the proportion of times that people preferred to receive a smell impulse along with an audio/visual stimulus for each of the three scenarios and across every budget. At small budget sizes, participants choose not to allocate resources for enabling the olfactory stimuli. As the budget increases, the frequency people select to receive an olfactory stimulus increases significantly. Furthermore, people tend to enable the control for smell more often in the Car and Bathroom VEs compared to the Kitchen VE. Participants’ preference to these scenarios might be explained due to the ease of visually locating the smell source in the scene.

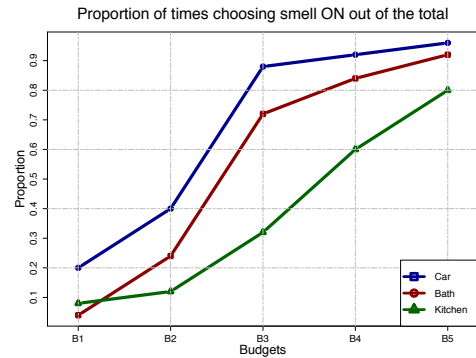


Fig. 7. Proportion of times smell stimulus was set on in each of the three scenarios and across all different budgets.

The main effect of budget was significant with  $F(4,96) = 61.47$ ,  $p < 0.05$  indicating a difference in the proportions the subjects selected to receive a smell impulse across the five budgets. The main effect of budget did not violate the assumption of sphericity (via Mauchly’s test,  $p > 0.05$ ). The main effect of scenario was also significant  $F(2,48) = 18.42$ ,  $p < 0.05$  indicating a difference in the proportions the smell burst was received by the participants across the three scenarios. The main effect for scenario did not also violate the assumption of sphericity. The interaction of budget  $\times$  scenario was also examined and it was not found to be statistically significant,  $F(8,192) = 1.76$ ,  $p > 0.05$ .

Contrast comparisons for proportions were conducted between groups of budgets using post-hoc proportion tests with Bonferroni corrections. These results are presented in Table 4. The results show that small budget sizes ( $B_1, B_2$ ) are not significantly different indicating that the budget size is still not sufficient for distributing resources to smell. Large budget sizes are also not significantly different ( $B_4, B_5$ ) meaning that people tend to enable the smell stimulus approximately the same number of times during their allocation.

Contrast comparisons for proportions were also conducted between groups of scenarios using post-hoc proportion tests and applying Bonferroni corrections. For  $B_1$  and  $B_2$  no significant changes were found.

Scenario	Budget size	p-value
Car	$(B_1, B_2)$ $(B_3, B_4, B_5)$	$< 0.05$
Bath	$(B_1, B_2)$ $(B_3, B_4, B_5)$	$< 0.05$
Kitchen	$(B_1, B_2)$ $(B_3, B_4, B_5)$	$< 0.05$
All	$(B_1, B_2)$ $(B_3, B_4, B_5)$	$< 0.05$

Table 4. Contrast comparisons for proportions of smell between budgets. Proportions of budgets with no significant differences are grouped together.

For the remaining three budget sizes, the group {Car, Bath} is significantly different from the group {Kitchen} indicating that participants clearly prefer to enable the smell on more frequently at scenarios of the former group.

#### 5.4.2 Analysis on the audio-visual preferences

Figure 8 depicts the mean percentages and confidence intervals for graphics and acoustics quality for each scenario and budget. As can be seen, the visual and auditory percentages are negatively correlated. For small budget sizes participants prioritise visuals and allocate the majority of the budget. As the budget size increases a balanced distribution of resources is preferred as the available amount permits it.

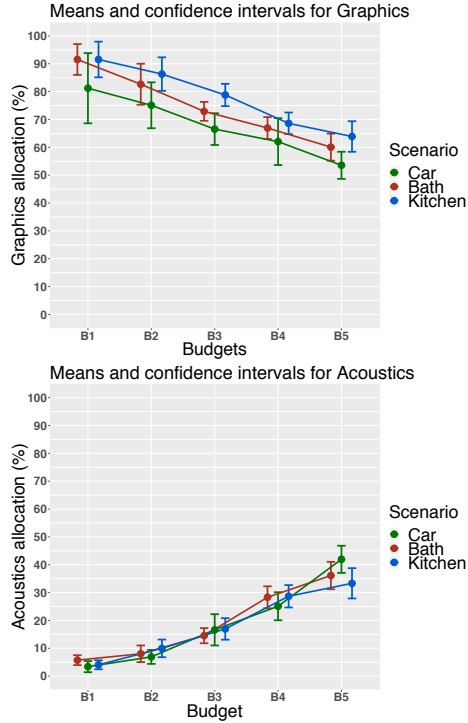


Fig. 8. Mean allocation percentages and confidence intervals for every scenario and across all the budget sizes Top: Graphics, Bottom: Acoustics. Jittering has been applied to the two plots for better visualisation.

MANOVA analysis was applied to examine the effect of the budget and scenario on the audio-visual allocation percentages. For the acoustic quality percentage, the main effect of budget was significant with  $F(4, 96) = 153.85$ ,  $p < 0.05$  while the effect of scenario was not found to be statistically significant with  $F(2, 48) = 0.0241$ ,  $p > 0.05$ . The main effect of budget and scenario did not violate the assumption of sphericity (via Mauchly's test,  $p > 0.05$ ). The interaction scenario  $\times$  budget was not found significant,  $F(8, 196) = 0.012$ ,  $p > 0.05$ .

For the visual quality percentage, the main effect of budget was significant with  $F(4, 96) = 42.11$ ,  $p < 0.05$ . The effect of scenario was also found to be statistically significant for the graphics allocation percentage with  $F(2, 48) = 11.03$ ,  $p < 0.05$ . The main effect of both

factors did not violate the assumption of sphericity (via Mauchly's test,  $p > 0.05$ ). The interaction scenario  $\times$  budget was not found significant,  $F(8, 196) = 0.94$ ,  $p > 0.05$ .

Contrast comparisons using post-hoc t-tests were also conducted to investigate groups of budgets that are not significantly different across the scenarios and also for finding groups of scenarios that are not significantly different across budgets. These contrasts are given in Table 5 and Table 6 for graphics and acoustics respectively.

Scenario	Budget size	p-value	Budget		Scenarios		p-value
			B1	B2	Car	Bath	
Car	$(B_1, B_2)$ $(B_3, B_4, B_5)$	$< 0.05$	B1	B2	Car	Bath	$< 0.05$
Bath	$(B_1, B_2)$ $(B_3, B_4, B_5)$	$< 0.05$	B1	B2	Car	Bath	$< 0.05$
Kitchen	$(B_1, B_2)$ $(B_3, B_4, B_5)$	$< 0.05$	B1	B2	Car	Bath	$< 0.05$
All	$(B_1, B_2)$ $(B_3, B_4, B_5)$	$< 0.05$	B1	B2	Car	Bath	$< 0.05$

Table 5. Left: Contrast comparisons between budgets at every scenario and across all the scenarios for graphics. Right: Contrast comparisons between scenarios at every budget and across all the budgets for graphics. Budgets or scenarios with no significant difference are grouped together.

As can be seen from Table 5, small budget sizes ( $B_1, B_2$ ) are not significantly different meaning that participants follow similar allocation strategy for graphics for these budgets independently of the scenario. The same argument also holds for very large budget sizes ( $B_4, B_5$ ). As far as the scenarios are concerned, subjects tend to increase graphics quality in a similar way for the Kitchen and Bath scenarios.

Scenario	Budget size	p-value	Budget		Scenarios		p-value
			B1	B2	Car	Kitchen	
Car	$(B_1, B_2)$ $(B_3, B_4, B_5)$	$< 0.05$	B1	B2	Car	Kitchen	$< 0.05$
Bath	$(B_1, B_2)$ $(B_3, B_4, B_5)$	$< 0.05$	B1	B2	Car	Kitchen	$< 0.05$
Kitchen	$(B_1, B_2)$ $(B_3, B_4, B_5)$	$< 0.05$	B1	B2	Car	Kitchen	$< 0.05$
All	$(B_1, B_2)$ $(B_3, B_4, B_5)$	$< 0.05$	B1	B2	Car	Kitchen	$< 0.05$

Table 6. Left: Contrast comparisons between budgets at every scenario and across all the scenarios for acoustics. Right: Contrast comparisons between scenarios at every budget and across all the budgets for acoustics. Budgets or scenarios with no significant difference are grouped together.

Table 6 shows that none of the scenarios is significantly different from the others when distributing resources for aural quality improvements. As far as the budgets are concerned, the same trend for graphics holds also for the acoustics for the effect of the budget size.

## 6 ESTIMATION MODEL

The experimental data was used in the construction of a statistical model which aims to estimate three quantities given a budget, and optionally the scenario. The first quantity is the probability that expresses whether smell should be given for an input budget and scenario while the other two estimations are proportions of the total budget that should be devoted to visual and aural quality.

As the decision to turn the smell stimulus on or off is based on a categorical variable, a logistic regression was used to model it. Specifically, if the probability to give a smell impulse is denoted as  $p$  then the ratio  $\frac{p}{1-p}$  expresses the odds of enabling the smell impulse release. This smell on/off decision can be modeled as:

$$\log\left(\frac{p}{1-p}\right) = \hat{\beta}_i^s + \hat{\beta}_b^s \cdot \text{budget} + \mathbb{1}_S \cdot \hat{\gamma}_S^s, \quad (8)$$

where the coefficients  $\hat{\beta}_i^s$ ,  $\hat{\beta}_b^s$  and  $\hat{\gamma}_S^s$  are the least squares regression estimates. The subscripts  $i$  and  $b$  are shorthand for the intercept and the budget respectively, and the subscript  $S$  is shorthand for scenario. The motivation for scenario specificity is dictated by the significant differences identified between the scenarios in the experimental data.  $\hat{\gamma}_S^s$  acts as a scenario specific offset to the intercept which can improve

the model for each scenario. The indicator function  $\mathbb{1}_S$  denotes whether the scenario is included as part of the model.

For an input budget size, if the signum of the right hand side of Equation 8 is positive then it can be inferred that  $p > 1 - p$  and thus a smell impulse should be delivered to the user. The above model is used to estimate whether the smell should be turned on/off and is only one component of the whole model. The full model is also composed of audio-visual percentage allocation components and can be written in matrix form as:

$$\begin{bmatrix} \hat{S}_1 \\ \hat{V}_1 \\ \hat{A}_1 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_i^s & \hat{\beta}_b^s & \hat{\gamma}_S^s \\ \hat{\beta}_i^v & \hat{\beta}_b^v & \hat{\gamma}_S^v \\ \hat{\beta}_i^a & \hat{\beta}_b^a & \hat{\gamma}_S^a \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \text{budget} \\ \mathbb{1}_S \end{bmatrix}, \quad (9)$$

where the letters  $\hat{V}_1$  and  $\hat{A}_1$  are the vision and audio allocation estimations while  $\hat{S}_1 = \log\left(\frac{p}{1-p}\right)$  is the model component for smell as given in formula 8.

The remainder of the paper will consider two forms of the model for both testing and evaluation:  $\mathbf{M}_1$  refers to  $\mathbb{1}_S = 0$  where the scenario is not considered, and  $\mathbf{M}_2$  refers to  $\mathbb{1}_S = 1$  which considers the scenario in the model.

For  $\mathbf{M}_1$ , hypothesis tests of the form:

$$H_0 : \beta_z^k = 0 \text{ vs } H_a : \beta_z^k \neq 0, \quad z \in \{i, b\} \text{ and } k \in \{s, v, a\}$$

were conducted to check the significance of the unknown parameters  $\beta_i^s, \beta_b^s, \beta_i^v, \beta_b^v, \beta_i^a, \beta_b^a$ . In all these tests the null hypothesis is rejected indicating that the respective least squares regression estimators ( $\hat{\beta}_i^s, \hat{\beta}_b^s, \hat{\beta}_i^v, \hat{\beta}_b^v, \hat{\beta}_i^a, \hat{\beta}_b^a$ ) should remain in the model formulation. The six estimates on the right hand side of the above equation are given in Table 7.

Model	$\hat{\beta}_i^s$	$\hat{\beta}_b^s$	$\hat{\beta}_i^v$	$\hat{\beta}_b^v$	$\hat{\beta}_i^a$	$\hat{\beta}_b^a$
$\mathbf{M}_1$	-1.31	0.03	84.89	-0.25	4.72	0.32

Table 7. Least squares regression estimates of the multivariate model  $\mathbf{M}_1$ . The subscripts “i” and “b” are used as shorthands for intercept and budget respectively. The superscripts “s”, “v”, “a”, are used instead of smell, vision and audio respectively.

The coefficients of determination  $R^2$  and  $R_{\text{adj}}^2$  for the goodness of fit for model  $\mathbf{M}_1$  are given in Table 8. The superscript in parenthesis is used to denote the sense. As the smell estimation is based on logistic regression the first coefficient of determination is based on the Akaike Information Criterion (AIC) (the lower the better) while the other two for graphics and audio are the Nagelkerke coefficients of determination (the higher the better).

Model	$R^{2(s)}$	$R^{2(v)}$	$R_{\text{adj}}^{2(v)}$	$R^{2(a)}$	$R_{\text{adj}}^{2(a)}$
$\mathbf{M}_1$	24.80	0.37	0.36	0.61	0.61
$\mathbf{M}_2$	18.40	0.40	0.40	0.65	0.62

Table 8. Coefficients of determination for the multivariate models  $\mathbf{M}_1$  and  $\mathbf{M}_2$ . The Akaike information criterion was used for the  $R^{2(s)}$ . The Nagelkerke coefficient is used for vision and audio.

$\mathbf{M}_2$  considers the scenario as input to the model, and therefore can better match the experimental results than  $\mathbf{M}_1$  which is based on the average across all scenarios.  $\mathbf{M}_2$  introduces extra parameters  $\gamma_S^s, \gamma_S^v$  and  $\gamma_S^a$  which are examined to see whether they are statistically significant and should be kept in the model  $\mathbf{M}_2$ . Hypothesis tests of the following form:

$$H_0 : \gamma_S^k = 0 \text{ vs } H_a : \gamma_S^k \neq 0, \quad k \in \{s, v, a\}, \quad (10)$$

were conducted to see the significance of these parameters in the Kitchen and Car scenario. For the Kitchen scenario ( $S = K$ ), it was

found that for the parameter  $\gamma_K^v$  the  $H_0$  is not rejected ( $p > 0.05$ ) and the same also holds for the parameter  $\gamma_K^a$ . For the car scenario,  $\gamma_C^a$  is also not rejected. The latter parameters were expected to be not significant as the factor *scenario* was not found to be statistically significant for adjusting aural quality in the MANOVA tests of the previous section. All the other parameters were found statistically significant (reject the  $H_0$ ).

The least square regression estimates of the  $\mathbf{M}_2$  model are given in Table 9, and the coefficients of determination  $R^2$  and  $R_{\text{adj}}^2$  for the goodness of fit for model  $\mathbf{M}_2$  are given in Table 8.

Sense	$\hat{\beta}_i$	$\hat{\beta}_b$	$\hat{\gamma}_C$ (Car)	$\hat{\gamma}_K$ (Kitchen)
Smell	-1.35	0.04	0.72	-1.10
Visual	87.33	-0.25	-7.29	*
Audio	4.72	0.32	*	*

Table 9. Least squares regression estimates of the multivariate model  $\mathbf{M}_2$ . The asterisks show the parameters that are not statistically significant (not reject the  $H_0$  hypothesis) and their estimators were excluded from the statistical model.

## 7 EXPERIMENT III: MODEL VALIDATION

The performance of the model was validated in an experimental study which included the use of untested budget sizes and a new untested scenario to test the performance of the model with inputs other than the ones used for the construction.

### 7.1 Design

The validation study follows the same guidelines as the main experimental study. Three scenarios were used, namely, Car, Kitchen and Kitti. Kitti was the new scenario for the validation stage (this was only used for training previously). Three budgets were used in this experiment, each one different from the previously used budgets. These were the midpoints between  $B_1$  and  $B_2$ ,  $B_2$  and  $B_3$  and  $B_3$  and  $B_4$ . These are denoted as  $NB_1$ ,  $NB_2$  and  $NB_3$ . All the materials of the main study, were also used for the validation stage with no change. The Kitchen scenario was used for training. A total of six people participated and none of these participants were familiar with the objectives of the experiment as they had not taken part in the previous experiments. We test both  $\mathbf{M}_1$  and  $\mathbf{M}_2$ , and for the scenario specific coefficients for  $\mathbf{M}_2$  we used the values from the Car scenario based on the rationale that the smell source has a similar configuration to the Car scenario.

### 7.2 Results

$\mathbf{M}_1$  and  $\mathbf{M}_2$  were compared with the data collected from the validation study to investigate their performance against real human preferences.

Figure 9 shows the means and confidence intervals for results for all scenarios and modalities. The model estimations are also depicted as colored curves for comparison. For graphics, the average error for absolute difference for the  $\mathbf{M}_1$  is 5.8% while for  $\mathbf{M}_2$  is 2.3%. For acoustics,  $\mathbf{M}_1$  and  $\mathbf{M}_2$  give an average error of absolute difference of 3.4%. For small budgets ( $NB_1$ ) both  $\mathbf{M}_1$  and  $\mathbf{M}_2$  predict equally well for graphics quality. As the budget size increases, model  $\mathbf{M}_2$  gives more accurate estimations.

For the Kitchen scene, for graphics, the average error for absolute difference is 4.4% and 1.6% for  $\mathbf{M}_1$  and  $\mathbf{M}_2$  respectively. For acoustics,  $\mathbf{M}_1$  and  $\mathbf{M}_2$  give an average error of absolute difference of 2.3%. For smell, at small budget sizes people prefer a low quality audio while smell is not set on. For the test budgets  $NB_1$  and  $NB_2$ ,  $\mathbf{M}_1$  and  $\mathbf{M}_2$  overestimate for audio and smell while they underestimate for graphics.

For the new Kitti scenario, whose data was *not* used to build  $\mathbf{M}_1$  or  $\mathbf{M}_2$ , for graphics the average error for absolute difference is 1.6% and 5.8% for  $\mathbf{M}_1$  and  $\mathbf{M}_2$  respectively while for acoustics  $\mathbf{M}_1$  and  $\mathbf{M}_2$  give an average error of absolute difference of 4.9% indicating that participants’ need for acoustic quality was significantly increased compared to the estimates given by either  $\mathbf{M}_1$  or  $\mathbf{M}_2$ .



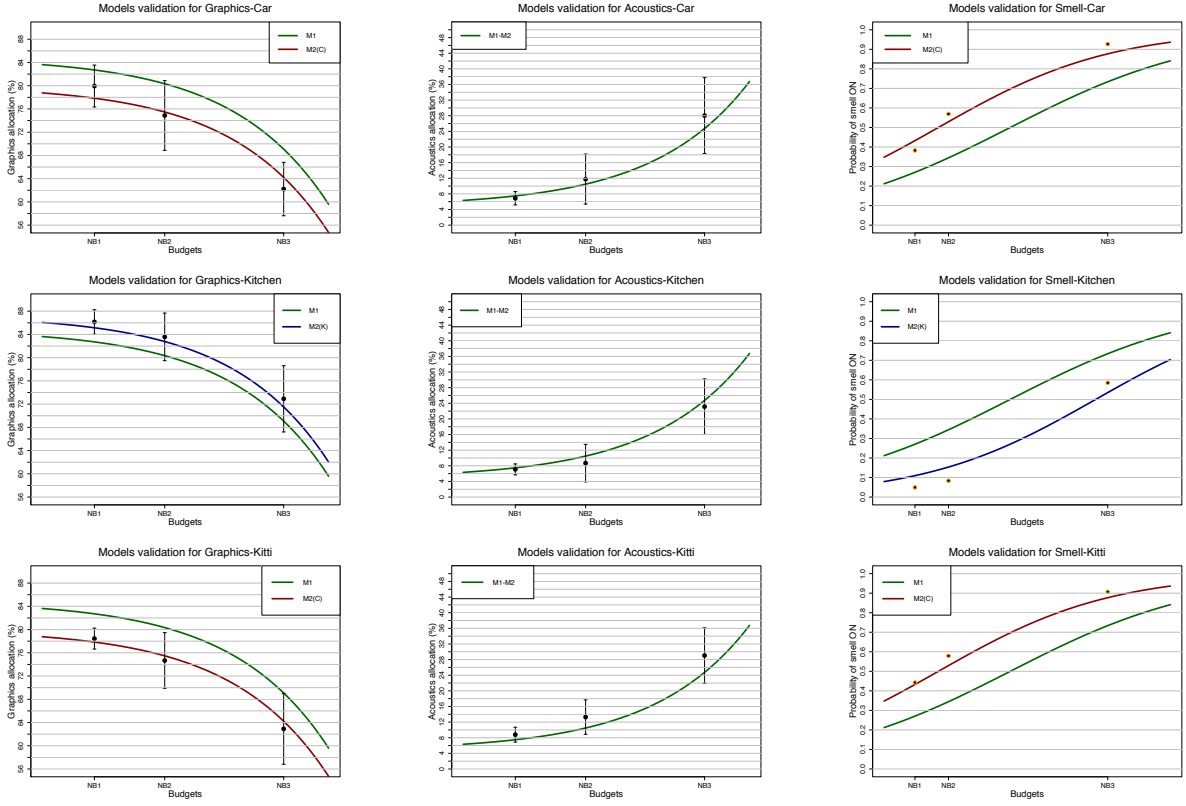


Fig. 9. Left and Middle: Means and confidence intervals for the three scenarios for graphics and acoustics respectively. The green curve shows  $M_1$  and the red shows  $M_2$ . The (K) or (C) notations of  $M_2$  denote coefficients from the Kitchen or Car scenario as used in  $M_2$ . Log-spacing was applied to the x-axis for better visualisation. Right: Proportion of times smell was preferred on for each of the test budgets across the three scenarios. The green and red curves are the probability estimations to turn the smell on for  $M_1$  and  $M_2$  respectively.

## 8 DISCUSSION

The results yield a number of potentially useful findings related to the way humans tend to allocate resources in tri-modal VEs. The co-existence of more than two senses in the experimental set-up does not affect humans' general trend to devote the majority of their budget for visual quality improvements. For the scenarios used and the smaller budget sizes, people's priority is to obtain a clear visual stimulus while compromises are made to both the audio quality and the addition of smell. As more resources become available, participants prefer an approximately balanced distribution of resources while the frequency with which smell was added was significantly increased.

As far as the effect of the scenario is concerned, it is clear that the four scenarios affect people's decision to enable the release of smell impulses. The scenario selection was also important for the visual allocation preferences while it had no effect on the percentage devoted to aural quality improvements. Participants did not show significant differences in acoustic resource allocation across the scenarios.

The validation stage shows evidence that the statistical model provides allocation estimates of different accuracy, especially in the case of the visual percentages.  $M_1$ , the generic version of the model, gives relatively accurate predictions but it was outperformed by model  $M_2$  in almost all the combinations of budget-scenario test conditions. The results of the validation study indicate also that both  $M_1$  and  $M_2$  give relatively accurate estimations for the three different validation budgets and scenario. When using  $M_2$ , results are improved at the cost of scene dependent parameters, which currently require experiments to determine. This is an area of future work, as it is likely that these parameters can be predicted as a function of the multi-modal content of the scene.

The delivery of the visual stimuli as static images is a limitation of this experimental study as it restricts the number of possible applications of the proposed estimation model. Both the auditory and

olfactory stimuli have an inherent temporal dimension (duration of the audio track/smell burst) that is not exploited with the use of static visual images. However, the experimental methodology can provide useful insights on extending to VEs with animations or interactive environments.

Another possible limitation is the use of the binary (on/off) smell stimuli. It can be said that the two-level methodology for delivering smell impulses restricts user's available options but it is not clear whether (and how) a hypothetical multiple quality scale for smell would benefit user's virtual experience, particularly with current olfactory hardware limitations, as shown in Experiment I. The existence of multiple smells, on the other hand, is considered more important in a multi-sensory VE and can admittedly increase the level of immersion as it is closer to what happens in the real world. To that end, the JND estimation methodology can be applied in arbitrary number of VOCs or combinations of them.

## 9 CONCLUSION

This paper presented a series of experimental studies with the ultimate goal of tri-modal resource allocation for VEs. Unlike previous work, which focused predominantly on audio-visual cross modalities, participants were given the opportunity to release an olfactory stimulus. Based on a first experiment, the sense of olfaction included two different quality levels (smell on/off) while smell impulses were delivered to the user using an olfactory display. This was used to develop a model for resource allocation for tri-modal virtual environments, which has been validated in a third experimental study. Future work will consider dynamic tri-modal VEs, where the temporal nature of visual, auditory and olfactory stimuli will be investigated. Additionally, we intend to explore whether the allocation of computational resources is impacted by user movement in the VE.

## REFERENCES

- [1] General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays. ITU-R, August 2012.
- [2] Pubchem compound database. Technical report, National Center for Biotechnology Information, 2016.
- [3] E. Acheson and M. Gamble. The applicability of weber's law to smell. *The American Journal of Psychology*, 10(1):82–142, 1898.
- [4] O. A. Ademoye, N. Murray, G.-M. Muntean, and G. Ghinea. Audio masking effect on inter-component skews in olfaction-enhanced multimedia presentations. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 12(4):51:1–51:14, Aug. 2016. doi: 10.1145/2957753
- [5] P. Bertelson and B. de Gelder. *Crossmodal Space and Crossmodal Attention*, chap. The Psychology of Multimodal Perception. Oxford University Press, USA, May 2004.
- [6] D. Burr and D. Alais. Combining visual and auditory information. *Prog Brain Res*, 155:243–258, 2006. doi: 10.1016/S0079-6123(06)55014-9
- [7] A. S. Azevedo and J. Jorge and P. Campos. Combining EEG Data with Place and Plausibility Responses as an Approach to Measuring Presence in Outdoor Virtual Environments. *Presence*, 23(4):354–368, Nov 2014.
- [8] F. Grani and S. Serafin and F. Argelaguet and V. Gouranton and M. Badawi and R. Gagne and A. Lecuyer. Audio-visual attractors for capturing attention to the screens when walking in CAVE systems. In *2014 IEEE VR Workshop: Sonic Interaction in Virtual Environments (SIVE)*, pp. 3–6, March 2014.
- [9] I. Bergström, S. Azevedo, P. Papiotis, N. Saldanha, and M. Slater. The plausibility of a string quartet performance in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1352–1359, April 2017. doi: 10.1109/TVCG.2017.2657138
- [10] Kajiya, James T. The rendering equation. *SIGGRAPH Computer Graphics*, 20(4):143–150, Aug. 1986.
- [11] M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, Jan 2002.
- [12] R. Skarbez and S. Neyret and F. P. Brooks and M. Slater and M. C. Whitton. A Psychophysical Experiment Regarding Components of the Plausibility Illusion. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1369–1378, April 2017.
- [13] Rocchesso, Davide. *Explorations in Sonic Interaction Design*. Logos, Verlag, Berlin, GmbH, 01 2011.
- [14] Slater, Mel and Spanlang, Bernhard and Corominas, David. Simulating Virtual Environments Within Virtual Environments As the Basis for a Psychophysics of Presence. *ACM Transactions Graphics*, 29(4):92:1–92:9, jul 2010.
- [15] K. Chen, B. Zhou, S. Chen, S. He, and W. Zhou. Olfaction spontaneously highlights visual saliency map. *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1768), 2013. doi: 10.1098/rspb.2013.1729
- [16] Y. Chen. Olfactory display: Development and application in virtual reality therapy. In *16th International Conference on Artificial Reality and Telexistence-Workshops (ICAT'06)*, pp. 580–584, Nov 2006.
- [17] A. Dhokia, E. Doukakis, A. Asadipour, C. Harvey, T. Bashford-Rogers, K. Debattista, B. Waterfield, and A. Chalmers. A calibrated olfactory display for high fidelity virtual environments. In *Proceedings of the Conference on Computer Graphics & Visual Computing, CGVC '16*, pp. 113–120. Eurographics Association, Goslar Germany, Germany, 2016.
- [18] H. Q. Dinh, N. Walker, L. F. Hodges, C. Song, and A. Kobayashi. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Proceedings of the 1999 IEEE Virtual Reality*, pp. 222–228. IEEE, March 1999.
- [19] E. Doukakis, K. Debattista, C. Harvey, T. Bashford-Rogers, and A. Chalmers. Audiovisual resource allocation for bimodal virtual environments. *Computer Graphics Forum*, 37(1):172–183, 2017.
- [20] J. Driver and C. Spence. Crossmodal attention. *Current Opinion in Neurobiology*, 8(2):245–253, Apr 1998.
- [21] N. I. Durlach and A. S. Mavor. *Virtual reality: Scientific and Technological Challenges*. Committee on Virtual Reality Research and Development, National Research Council report, 1995.
- [22] G. Ghinea, C. Timmerer, W. Lin, and S. R. Gulliver. Mulsemmedia: State of the art, perspectives, and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 11(1s):17:1–17:23, Oct. 2014. doi: 10.1145/2617994
- [23] W. Haag and C. Wrenn. *The PID Handbook-Theory and Applications of Direct-Reading Photoionization Detectors 2nd Edition*. CA: RAE Systems Inc., 2006.
- [24] D. Harel, L. Carmel, and D. Lancet. Towards an odor communication system. *Computational Biology and Chemistry*, 27(2):121–133, 2003.
- [25] C. Harvey, T. Bashford-Rogers, K. Debattista, E. Doukakis, and A. Chalmers. Olfaction and selective rendering. *Computer Graphics Forum*, 37(1):350–362, 2018.
- [26] C. Harvey, K. Debattista, T. Bashford-Rogers, and A. Chalmers. Multimodal perception for selective rendering. *Computer Graphics Forum*, 36(1):172–183, 2016.
- [27] C. Hendrix and W. Barfield. The sense of presence within auditory virtual environments. *Presence: Teleoperators and Virtual Environments*, 5(3):290–301, 1996. doi: 10.1162/pres.1996.5.3.290
- [28] R. S. Herz and T. Engen. Odor memory: Review and analysis. *Psychonomic Bulletin & Review*, 3(3):300–313, 1996.
- [29] V. Hulusic, C. Harvey, K. Debattista, N. Tsingos, S. Walker, D. Howard, and A. Chalmers. Acoustic rendering and auditory-visual cross-modal perception and interaction. *Computer Graphics Forum*, 31(1):102–131, Feb. 2012.
- [30] F. Kingdom and N. Prins. *Psychophysics*. Academic Press, 2009.
- [31] B. Kollmeier, T. Brand, and B. Meyer. *Springer Handbook of Speech Processing*, chap. Perception of Speech and Sound, pp. 61–82. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
- [32] P. Larsson, D. Vastfjall, and M. Kleiner. Better presence and performance in virtual environments by improved binaural sound rendering. In *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*, Jun 2002.
- [33] T. R. Marrero and E. A. Mason. Gaseous diffusion coefficients. *Journal of Physical and Chemical Reference Data*, 1(1):3–118, 1972.
- [34] T. Moeck, N. Bonneel, N. Tsingos, G. Drettakis, I. Viaud-Delmon, and D. Alloza. Progressive perceptual audio rendering of complex scenes. In *I3D '07: Proceedings of the 2007 symposium on Interactive 3D graphics and games*, pp. 189–196. ACM, New York, NY, USA, 2007. doi: 10.1145/1230100.1230133
- [35] B. G. Munyan, S. M. Neer, D. C. Beidel, and F. Jentsch. Olfactory stimuli increase presence in virtual environments. *PLoS ONE*, 11(6), 2016. doi: 10.1371/journal.pone.0157568
- [36] T. Nakamoto and K. Yoshikawa. Movie with scents generated by olfactory display using solenoid valves. In *IEEE Virtual Reality Conference (VR 2006)*, pp. 291–292, March 2006.
- [37] M. Narwaria, R. Mantiuk, M. Perreira Da Silva, and P. Le Callet. HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images: a calibrated method for objective quality prediction of high-dynamic range and standard images. *Journal of Electronic Imaging*, 24(1):010501, Dec. 2014.
- [38] A. J. Nicell. Assessment and regulation of odour impacts. *Atmospheric Environment*, 43(1):196–206, 2009.
- [39] P. Sande and S. Ray. Mesh size effect on CFD simulation of gas-fluidized geldart a particles. *Powder Technology*, 264:43–53, 2014.
- [40] S. S. Schiffman and T. C. Pearce. *Introduction to Olfaction: Perception, Anatomy, Physiology, and Molecular Biology*, pp. 1–31. Wiley-VCH Verlag GmbH & Co. KGaA, 2004.
- [41] A. Seigneure, K. Durand, T. Jiang, J. Baudouin, and B. Schaal. The nose tells it to the eyes: crossmodal associations between olfaction and vision. *Perception*, 39(10):1541 – 1554, 2010.
- [42] H.-S. Seo, E. Roidl, F. Miller, and S. Negoias. Odors enhance visual attention to congruent objects. *Appetite*, 54(3):544 – 549, Jun 2010. Epub.
- [43] S. Siltanen, L. Tapio, S. Kiminki, and L. Savioja. The room acoustic rendering equation. *Acoustical Society of America*, 122(3):1624–1632, 2007.
- [44] R. L. Storms. *Auditory-Visual Cross-Modal Perception Phenomena*. {PhD} thesis, Naval Postgraduate School, 1998.
- [45] S. Tsan-Hsing, W. William, S. Aamir, Y. Zhigang, and Z. Jiang. A new k-ε eddy viscosity model for high reynolds number turbulent flows. *Computers & Fluids*, 24(3):227 – 238, 1995.